MycoKeys 86: 195–212 (2022)
doi: 10.3897/mycokeys.86.77431
https://mycokeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
MycoKeys
Launched to accelerate biodiversity research

A long-read amplicon approach to scaling up the metabarcoding of lichen herbarium specimens

Cécile Gueidan[1], Lan Li[1]

1 *Australian National Herbarium, National Research Collections Australia, NCMI, CSIRO, Canberra, ACT, 2601, Australia*

Corresponding author: Cécile Gueidan (cecile.gueidan@csiro.au)

Academic editor: Imke Schmitt | Received 1 November 2021 | Accepted 24 January 2022 | Published 2 February 2022

Citation: Gueidan C, Li L (2022) A long-read amplicon approach to scaling up the metabarcoding of lichen herbarium specimens. MycoKeys 86: 195–212. https://doi.org/10.3897/mycokeys.86.77431

Abstract

Reference sequence databases are critical to the accurate detection and identification of fungi in the environment. As repositories of large numbers of well-curated specimens, herbaria and fungal culture collections have the material resources to generate sequence data for large number of taxa, and could therefore allow filling taxonomic gaps often present in reference sequence databases. Financial resources to do that are however often lacking, so that recent efforts have focused on decreasing sequencing cost by increasing the number of multiplexed samples per sequencing run while maintaining high sequence quality. Following a previous study that aimed at decreasing sequencing cost for lichen specimens by generating fungal ITS barcodes for 96 specimens using PacBio amplicon sequencing, we present a method that further decreases lichen specimen metabarcoding costs. A total of 384 mixed DNA extracts obtained from lichen herbarium specimens, mostly from the four genera *Buellia*, *Catillaria*, *Endocarpon* and *Parmotrema*, were used to generate new fungal ITS sequences using a Sequel I sequencing platform and the PacBio M13 barcoded primers. The average success rate across all taxa was high (86.5%), with particularly high rates for the crustose saxicolous taxa (*Buellia*, *Catillaria* and others; 93.3%) and the terricolous squamulose taxa (*Endocarpon* and others; 96.5%). On the other hand, the success rate for the foliose genus *Parmotrema* was lower (60.4%). With this taxon sampling, greater specimen age did not appear to impact sequencing success. In fact, the 1966–1980 collection date category showed the highest success rate (97.3%). Compared to the previous study, the abundance-based sequence denoising method showed some limitations, but the cost of generating ITS barcodes was further decreased thanks to the higher multiplexing level. In addition to contributing new ITS barcodes for specimens of four interesting lichen genera, this study further highlights the potential and challenges of using new sequencing technologies on collection specimens to generate DNA sequences for reference databases.

Keywords

Collection specimens, ITS barcode, lichenised fungi, PacBio amplicon sequencing

Copyright *Cécile Gueidan & Lan Li*. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

Reference nucleotide sequence databases aim at providing access to curated and high-quality nucleotide sequences representing a broad taxonomic range of living organisms. They are critical to the accurate detection and identification of organisms from environmental samples and, for organisms lacking diagnostic characters, they are a useful tool to confirm morphology-based identifications. In fungi, the internal transcribed spacer region (ITS) has historically been used for species-level molecular identification (Gardes and Bruns 1993; Kõljalg et al. 2005; Abarenkov et al. 2010) and this gene region was later chosen as the fungal universal barcode (Schoch et al. 2012). Well-curated and high-quality fungal ITS sequences are now available from several databases, including RefSeq (Schoch et al. 2014; O'Leary et al. 2015), UNITE (Kõljalg et al. 2013; Nilsson et al. 2018) and ISHAM-ITS (Irinyi et al. 2015). However, the taxonomic coverage for fungi represented in these sequence databases remains incomplete (Orok et al. 2012; Kõljalg et al. 2013; Crous et al. 2014; Nilsson et al. 2018). As well-curated resources of dried or living material of large numbers of fungal species, herbaria and culture collections can contribute to fill some of the taxonomic gaps in these sequence databases (Yahr et al. 2016; Gueidan et al. 2019), and guarantee high taxonomic standards, in particular by prioritising the sequencing of generic and species types (Crous et al. 2014).

Taking advantage of the development of next generation sequencing (NGS) methods, large numbers of fungal ITS sequences have been generated these last ten years. Fungal metabarcoding studies that detect and identify fungi in environmental samples based on inferred operational taxonomic units have mostly generated partial ITS sequences, either ITS1 or ITS2 (Nilsson et al. 2010; Mello et al. 2011; Blaalid et al. 2013). This stems from their preferential use of Illumina sequencing technology which, although allowing affordable high-quality mass molecular barcoding, restricts the maximum read length to 300 bp. Other fungal metabarcoding studies have used long-read technologies, either Roche 454 pyrosequencing (Buée et al. 2009; Lumini et al. 2010; Blaalid et al. 2012; Geml et al. 2014), Pacific Biosciences SMRT sequencing (Chen et al. 2015; Cline and Zak 2015; James et al. 2016; Schlaeppi et al. 2016; Walder et al. 2017; Heeger et al. 2018; Terdersoo and Anslan 2019; Castaño et al. 2020), or Oxford Nanopore MinION sequencing (Hu et al. 2019; Loit et al. 2019) to generate ITS sequences or other molecular barcodes. For lichenised fungi, early long-read metabarcoding studies used Roche 454 pyrosequencing technology (Hodkinson and Lendemer 2013; Lücking et al. 2014; Mark et al. 2016). After the decline of this technology, Pacific Biosciences SMRT sequencing was shown to be a viable option to generate full length high-quality sequences from lichen herbarium specimens (Gueidan et al. 2019).

Although used for whole genome sequencing of lichen metagenomes (Tzovaras et al. 2020), to our knowledge, PacBio SMRT sequencing has only been used for metabarcoding purposes in one study involving lichenised fungi (Gueidan et al. 2019). In this previous study, ITS sequences of 96 lichen specimens were amplified using a

two-step PCR approach, with modified ITS primers and PacBio barcoded universal primers. PCR products were then sequenced using the PacBio RS II platform and assembled and denoised using Long Amplicon Analysis (LAA; Bowman et al. 2014). High quality ITS sequences were generated for 88.5% of the samples, with a cost per sample of AU$37. The mixed DNA samples resulting from the DNA extractions of these 96 lichen herbarium specimens also allowed the sequencing of other associated fungi. Here, the same method is used to generate ITS sequences from 384 lichen herbarium specimens, with the main goal of further decreasing the cost per sample. A new set of barcoded universal primers (M13 barcoding system, Larrea et al. 2018) developed by PacBio (Menlo Park, CA, USA) was tested in this study, as well as the then new PacBio Sequel I sequencing platform. Finally, sequence assembly and denoising were performed with a different pipeline, which included SMRT Tools (PacBio, Menlo Park, CA, USA) and DADA2 (Callahan et al. 2016).

The main goals of this study were to 1) assess the current cost and efficiency of a PacBio metabarcoding method applied to lichen herbarium specimens following changes in laboratory and bioinformatic pipelines, and 2) generate high-quality ITS sequences to contribute to reference sequence databases, as well as to molecular taxonomic studies of several lichen groups.

Material and methods

Taxon sampling and DNA extractions

For this study, 384 lichen specimens were selected because of their importance to several ongoing taxonomic works on Australian lichens at the Australian National Herbarium (see Suppl. material 1: Table S1). Four main genera were represented (Fig. 1A–D): *Catillaria* (14 specimens from Australia and France), *Buellia* (99 specimens from Australia), *Endocarpon* (167 specimens, including 157 from Australia) and *Parmotrema* (96 specimens from Australia). Additionally, a few other specimens were sampled: *Sporastatia* (4 specimens), and *Halecania* (1 specimen) and 3 unidentified specimens (one crustose saxicolous species and two squamulose terricolous species). The majority of the specimens were identified to the species or genus levels. The specimens were collected between 1966 and 2018, and are kept at CANB, UNSW, NSW, MARSSJ, ABL, BM, HO and in the private herbaria of P. McCarthy, M. Bertrand, B. McCune, and D. Stone. The material of crustose specimens (eg, *Catillaria*, *Buellia*) was detached from the substrate with a clean single-edge razor blade and a weigh paper was used to collect and transfer it to tubes containing a banded ceramic sphere and garnets (2 mL Lysing Matrix A, MP Biomedicals, Seven Hills, NSW, Australia). For squamulose and foliose species (e.g., *Endocarpon*, *Parmotrema*), lobes or squamules were detached from the substrate using clean tweezers and transferred directly to Lysing Matrix A tubes.

The samples were ground with a Precellys Evolution (Bertin Instruments, Montigny-le-Bretonneux, France) in 2–3 cycles of 30 sec at 6,000 rpm. To avoid cross-con-

Figure 1. Examples of lichen herbarium specimens used for this study **A** *Parmotrema perlatum*, specimen J.A. Elix 43686 (CANB790817) **B** *Endocarpon pusillum*, specimen H. Streiman 45100 (CBG9011273) **C** *Buellia albula*, specimen J.A. Elix 45138 (CANB810791) **D** *Catillaria* sp., specimen J.A Elix 37142 (CANB872684). Scale bar: 1 cm. Photos C. Gueidan.

taminations, the tubes were briefly centrifuged before the caps were removed. Genomic DNA was extracted using the Invisorb DNA Plant HTS 96 kit (Stratec Molecular, Berlin, Germany) adhering to the manufacturer's instructions, except for the few following modifications. The lysis buffer and proteinase K were added to each tube of ground material, which were then manually homogenised and incubated at 65 °C for 1 hour. The tubes were centrifuged at 11,000 rpm for 2 min and the supernatants were transferred onto the 96-well prefilter plate using a width-adjustable multichannel pipette. The RNase A (40 µl/well of a 10 mg/ml solution) was added after the prefiltration step and the tubes were incubated at room temperature for 15–20 min before adding the binding buffer. The last centrifugation step was changed to 10 min at 2,000 rpm (instead of 5 min at 4,000 rpm) to avoid breaking the elution plates. The DNA was eluted in 100 µl of elution buffer and 1/10 dilutions of the DNA samples were prepared.

Amplification, normalisation and pooling

Indexed PCR products were generated using a 2-step PCR approach as described in the PacBio Barcoded Universal Primers protocol (https://www.pacb.com/wp-content/

uploads/2015/09/Procedure-and-Checklist-Preparing-SMRTbell-Libraries-PacB-Barcoded-Universal-Primers.pdf), but with few modifications. The fungal ITS barcode (internal transcribed spacer 1, 5.8S ribosomal RNA subunit and internal transcribed spacer 2) was the target region. With a first PCR, our target region was amplified using the primers ITS1F (Gardes and Bruns 1993) and ITS4 (White et al. 1990), both modified by adding a 5' block and a tail representing the PacBio M13 primer sequence (Larrea et al. 2018): /5AmMC6/GTA AAA CGA CGG CCA GTC TTG GTC ATT TAG AGG AAG TAA for ITS1F-M13 and /5AmMC6/CAG GAA ACA GCT ATG ACT CCT CCG CTT ATT GAT ATG C for ITS4-M13. In a 25 µl reaction, 5 µl of MyFi buffer (Bioline, London, UK), 1 µl of MyFi polymerase, 2.5 µl of 3 µM of each primer, 13 µl of water and 1 µl of DNA template were added. For each plate, the amplification was done twice, once using the raw DNA extracts and once using a 1/10 dilution of the raw DNA extracts. The 96 PCR reactions were performed in strip tubes with individual caps to avoid cross-contaminations. The PCR program was 5 min at 95 °C, then 20–25 (and occasionally 35) cycles of 30 sec at 95 °C, 30 sec at 53 °C and 1:30 min at 72 °C, followed by a final elongation step for 7 min at 72 °C. Selected PCR products were run onto an agarose gel using the nucleic acid stain GelRed (Biotium, Fremont, CA, USA). If the gel did not show primer dimer bands, no PCR product cleaning was undergone at this stage.

A second amplification was then performed using part of a set of 64 barcoded M13 primers (32 forward and 32 reverse) provided by PacBio (Menlo Park, CA, USA). The barcode sequences were 16 bp long (see Suppl. material 1: Table S2 for their names and sequences). For the second amplification, the PCR products resulting from the raw extracts and the ones resulting from the 1/10 dilutions were first pooled together in order to minimize the number of negative samples. In a 25 µl reaction, 5 µl of MyFi buffer, 1 µl of MyFi polymerase, 2.5 µl of each barcoded primer pairs (3 µM), 13 µl of water and 1 µl of the pooled product of the first round of PCRs were added. The PCR program was 5 min at 95 °C, then 25 cycles of 30 sec at 95 °C, 30 sec at 65 °C and 1:30 min at 72 °C, followed by a final elongation step for 5 min at 72 °C. All PCR products were checked on a gel as previously described and cleaned using AMPureXP beads (Beckman Coulter, Brea, CA, USA). First, the beads were prewashed as recommended for SMRTbell™ library preparation by the Ramaciotti Centre for Genomics (https://www.ramaciotti.unsw.edu.au/sites/default/files/2019-04/RAMAC_Long_Linked_Read_guidelines_2019.pdf). The cleaning was then done by adding 0.8X volume of beads to each PCR product, followed by two washes with 200 µl of 70% ethanol. Dry beads were then resuspended in 25 µl of the EB elution buffer (Qiagen, Hilden, Germany). The concentrations were measured with a Nanodrop 8000 spectrophotometer (Thermo Scientific, Waltham, MA, USA) and 10 ng/µl working solutions were manually prepared. One µl of each of the 384 samples was then pooled into a single tube.

Library preparation, sequencing and primary analysis

The pooled sample was sent to the Ramaciotti Centre for Genomics (UNSW Sydney, Australia) for single molecular real-time (SMRT) sequencing. The library preparation

was done using the SMRTbell Template Prep Kit v. 1.0 (Pacific Biosciences, Menlo Park, CA, USA). The sample was sequenced in one SMRT cell and with a ten-hour movie, using the Sequel Binding Kit v. 3.0 and the Sequel Sequencing Plate v. 3.0 (Pacific Biosciences). The subread bam file provided by Ramaciotti was generated using SMRT Link v. 6.0 (Pacific Biosciences). This subread bam file was then demultiplexed using the "lima" command in SMRT Tools v. 7.0.1 (Pacific Biosciences), and the circular consensus sequences (CCSs) generated using the "ccs" command (0.9999 minimum predicted accuracy and 3 minimum passes).

Secondary analysis and sequence identification

Generated CCSs were denoised using DADA2 v. 1.14 (Callahan et al. 2016), a software that infers sequence variants from high-throughput amplicon sequencing datasets. A custom R script, written by the bioinformatics team at the BRF (ANU, Canberra, Australia) and available upon request, allowed the batch processing of each CCS fastq file through the DADA2 pipeline. First, the primers were removed and the reads reoriented. Read quality profiles and length distribution histograms were then generated for each file. The reads were filtered (maxEE = 1, minQ = 3 and minLen = 300) and dereplicated. The error model was then estimated from the data (BAND-SIZE = 32) and the data further filtered for errors. After a last check for chimeras, fasta files of sequence variants were generated. For sequence identification, a blastn query (BLASTN 2.9.0+) was performed on these files against the NCBI nt database using a max_target_seqs of 3.

The blastn output was parsed into a single text file using a custom script and the results checked manually. Sequencing was considered successful if one of the generated sequence variants matched the same genus as the target taxa. For the unsuccessful samples, the fastq ccs files were converted to fasta using the fastqtofasta command in fastx 0.0.14, and an additional blastn query (BLASTN 2.12.0+) was performed on the ccs files using the same parameter as above. The blastn results were checked manually and sequencing was considered successful if one of the ccs matched the same genus as the target taxa. Demultiplexed fastq files were deposited in the Sequence Read Archive on NCBI (BioProject ID PRJNA796455).

Results

Amplification and sequencing

The two-step amplification approach generated PCR products with concentrations ranging from 11 to 1,573 ng/µl (Suppl. material 1: Table S1). The average concentration was 90 ng/µl and only 51 of the 384 samples were under 50 ng/µl (Suppl. material 1: Table S1). A total of 1.16 µg of PCR products was submitted to Ramaciotti in a pooled sample. The sample met the quality control requirements and showed DNA fragments ranging from 581 to 1,447 bp, with two clear peaks at 694 and 998 bp,

which are within the expected size range for the ITS barcode. The SMART cell generated 217,195 polymerase reads, with a mean length of 52,801 bp. This corresponded to 12,955,790 subreads with a mean length of 841 bp, and an average of 60 passes per ccs.

Sequence analysis and sequence identity

Using SMRT Tools, CCSs were recovered for 372 of the 384 samples, with only 12 samples for which no reads were generated (Suppl. material 1: Table S1). For each positive sample, between 22 and 1,368 CCSs were generated. After sequence denoising using the DADA2 amplicon pipeline, 1 to 21 sequence variants were found per sample. A blast analysis was conducted on all sequence variants and the results were compared to the morphology-based genus and/or species identification. Following this denoising step, a sequence of the target taxa was generated for 262 of the 384 initial samples. For 70 samples, no sequence variant matching the target taxon was generated by DADA2, but one or more sequences of the target taxon could be recovered from the CCS files. Therefore, in total, 332 of the 384 initial samples were considered as successful (86.5%). For 40 samples, sequence of the target taxon could not be found neither amongst the sequence variants generated by DADA2, nor among the CCSs. The majority of the samples that failed to generate sequences from the target species were from specimens of the genus *Parmotrema* (33 samples out of 40, or 82.5%).

When divided into three main morphological groups of taxa (Fig. 2), *Buellia*, *Catillaria* and other saxicolous crustose taxa had a sequencing success rate of 93.3% (111 positive samples out of 119). *Endocarpon* and other terricolous squamulose taxa had a sequencing success rate of 96.5% (163 positive samples out of 169 sample). Finally, the foliose genus *Parmotrema* had a sequencing success rate of only 60.4% (58 positive samples out of 96). When divided in five categories of time of collection (1966–1980, 1981–1990, 1991–2000, 2001–2010 and 2011–2020), the proportion of unsuccessful samples does not increase with the age of specimens: all but one category had more than 90% success rate (Fig. 3). In fact, the highest success rate was for the oldest class of specimens, 1966–1980, with 97.2%. The lower success rate for the 2001–2010 class (64.5%) was due to the lower success rate of specimens of *Parmotrema*, most of which were collected between 2005 and 2010.

Discussion

Building upon a previous work (Gueidan et al. 2019), the goal of this study was to generate high-quality ITS sequences for 384 lichen herbarium specimens and assess the current efficiency and cost of a modified PacBio metabarcoding method that had previously been applied to lichen herbarium specimens.

Generation of ITS barcodes for 384 lichen herbarium specimens

ITS sequences were successfully generated for 332 of the 384 herbarium specimens included in this study. Most of the specimens included belonged to the four genera

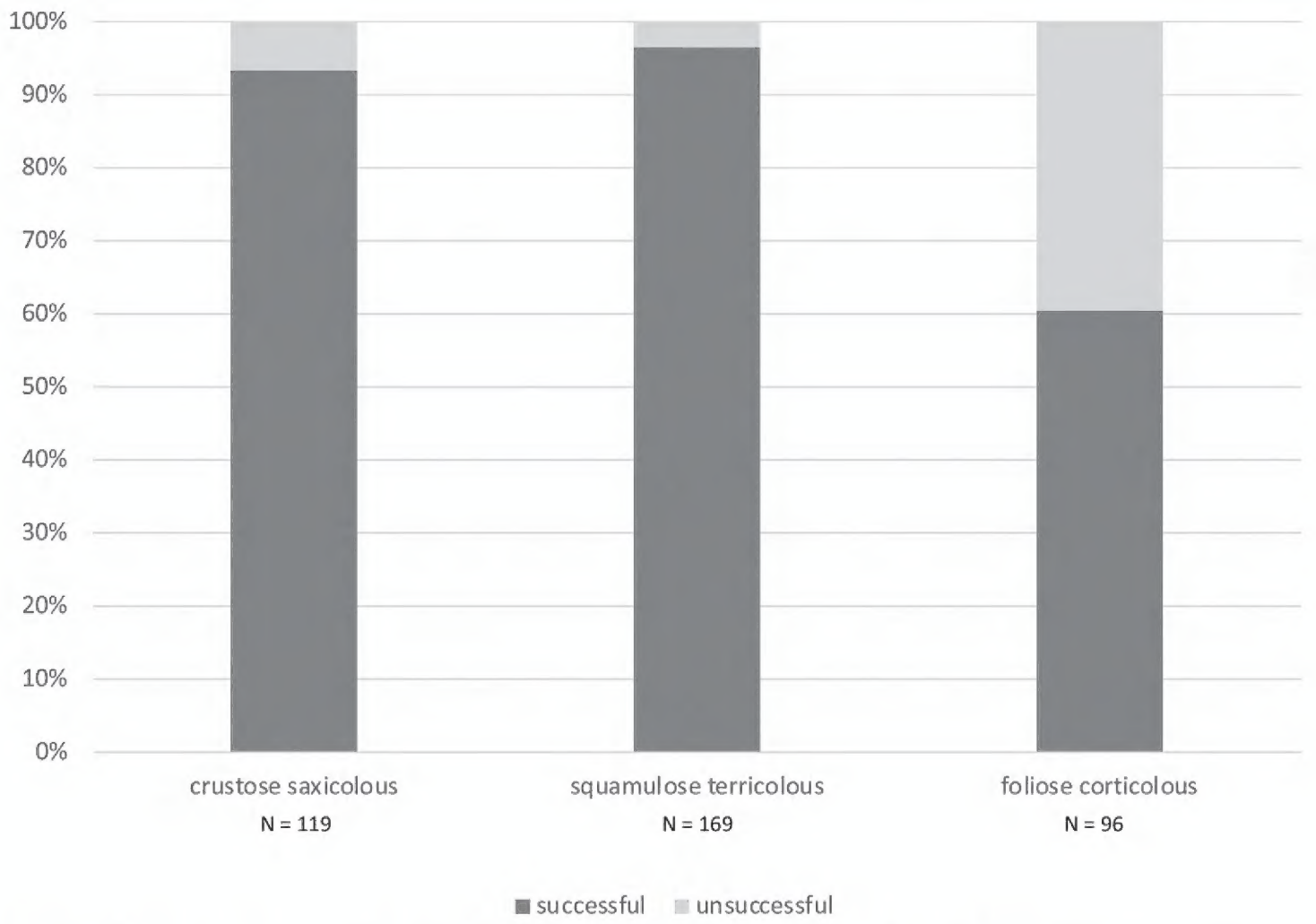

Figure 2. Sequencing success for different morphological groups of taxa included in this study. Specimens were grouped into three main morphological categories: **1** *Buellia*, *Catillaria* and other crustose saxicolous taxa **2** *Endocarpon* and other squamulose terricolous taxa **3** the foliose corticolous genus *Parmotrema*. In the graph, stalked columns show successful samples (sequence generated for the target species) in dark grey and unsuccessful samples (no sequence generated or generated sequences not from the target species) in light grey. The total number of samples (N) is indicated below each corresponding column.

Buellia, *Catillaria*, *Endocarpon* and *Parmotrema*. The success rate for the sequencing of the target ITS barcode was high (an average of 86.5% across all taxa) and similar to the one reported in Gueidan et al. (2019), which was of 88.5%. The sequencing success rate for the genus *Parmotrema* was, however, particularly low (60.4%), especially in comparison to *Buellia*, *Catillaria* and other crustose taxa (93.3%) and *Endocarpon* and other squamulose taxa (96.5%). Although a PCR product was obtained for all the *Parmotrema* samples, generated ITS sequences did often not belong to the target species, but to other fungi. This is likely due to inefficient amplification of the ITS barcode from the target *Parmotrema* species using the ITS1F-ITS4 primer pair. Although this primer pair is commonly used to amplify ITS of lichenised fungi (e.g., James et al. 2006; Kelly et al. 2011; Mark et al. 2016), it seems to have been used in combination with one other primer pair, ITS1-LM (Myllys et al. 1999) and ITS2-KL (Lohtander et al. 1998), in previous studies on the genus *Parmotrema* (Divakar et al. 2005; Del Prado et al. 2011). This indicates a possible amplification issue with the ITS1F-ITS4 primer pair for these species. Further sequencing will be carried out in the future for this genus using the alternative primer pair ITS1-LM/IT2-KL, to attempt recovering ITS barcodes for additional *Parmotrema* species.

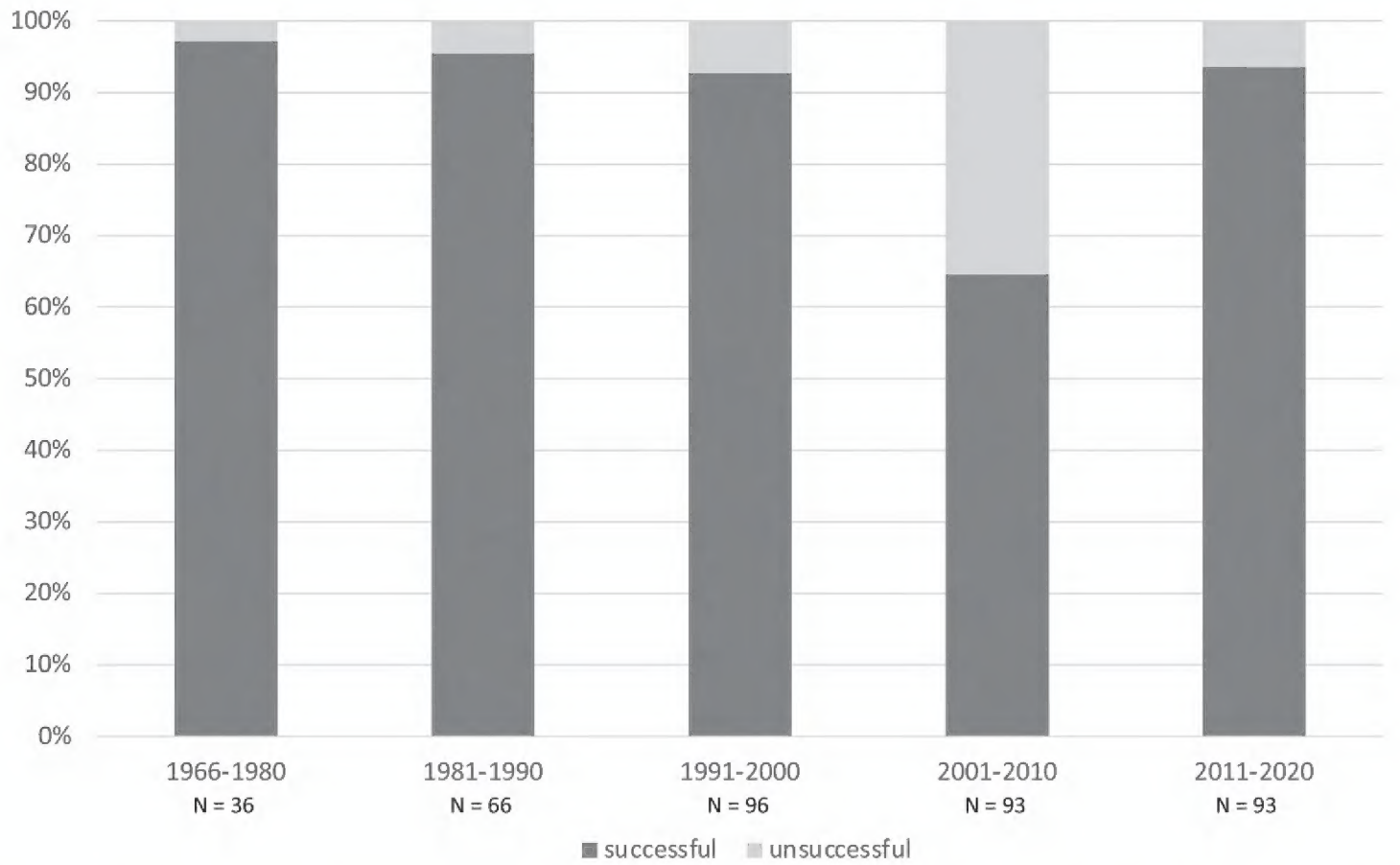

Figure 3. Sequencing success for different ages of specimens included in this study. Specimens were grouped in five categories: 1966–1980, 1981–1990, 1991–2000, 2001–2010, 2011–2020. In the graph, stalked columns show successful samples (sequence generated for the target species) in dark grey and unsuccessful samples (no sequence generated or generated sequences not from the target species) in light grey. The total number of samples (N) is indicated below each corresponding column.

The genomic DNA of some groups of lichens, most often from crustose corticolous tropical families (e.g., Staiger et al. 2006 for the Graphidaceae; Weerakoon et al. 2012 and Gueidan et al. 2016 for the Pyrenulaceae), is notoriously difficult to obtain from dried specimens. However, for most lichens, ITS sequences can usually easily be obtained from extracts of a large range of lichen herbarium specimens, including some relatively old ones. The oldest ones were a 75-year-old specimen of *Aspicilia aschabadensis* (Sohrabi et al. 2010), a 100-year-old specimen of *Staurolemma omphalarioides* (Bendiksby et al. 2014), and a 151-year-old specimen of *Caloplaca conversa* (Redchenko et al. 2012). A sequence of the small subunit of the mitochondrial ribosomal RNA gene (mtSSU) was also obtained from a 127-year-old specimen of *Peltigera collina* (Kistenich et al. 2019). In the latter study, which used Ion Torrent sequencing to generate mtSSU sequences from historical lichen specimens collected from 1885 onwards, specimen age had a significant influence on sequencing success, with older specimens less likely to yield good quality sequences (Kistenich et al. 2019). In our study, ages of lichen herbarium specimens ranged from 2 to 54 years. Apart from the specimen age category which included most *Parmotrema* samples (2001–2010), for which a primer issue caused low sequencing success, all age categories had high sequencing success rates, ranging from 92.7% to 97.2% (Fig. 3). Therefore, at least for some of the represented groups (*Endocarpon*

and saxicolous *Buellia* and *Catillaria*), generating ITS sequences for specimens up to 50 years old was not an issue. These results highlight the importance and relevance of older herbarium specimens, including types, in molecular taxonomy.

Efficiency and cost of the applied method

PacBio long read sequencing is a powerful approach, which when applied to amplicons, can utilise circular sequencing to generate high quality consensus of shorter nucleotide fragments. In order to correct sequencing errors, subreads extracted from one polymerase read – therefore generated from a single amplicon molecule, are aligned and assembled into one circular consensus sequence (CCS). Following CCS generation, additional software and pipelines are available to further correct sequencing errors, a step often called denoising. Although several software are available for denoising Illumina amplicon data (e.g., unoise, Edgar 2016; deblur, Amir et al. 2017), very few are available for PacBio amplicon data. In a previous study (Gueidan et al. 2019), a denoising approach developed for allele phasing of PacBio amplicon data (LAA) was used to generate high quality ITS sequences from lichen herbarium specimens. Parallel Sanger sequencing for a subset of samples confirmed that this method worked well for generating high-quality sequences for target and associated fungal species (Gueidan et al. 2019). However, a more relevant approach for metabarcoding is available (DADA2, Callahan et al. 2016) and was tested here. DADA2 is a software package that allows to generate amplicon sequence variants from high-throughput amplicon sequencing data. Conveniently, it can use PacBio data and a protocol is available for the size-variable ITS marker.

Despite DADA2 generating target sequence variants for a large number of our samples, a significant number of samples (70) did not yield sequences from the target taxon despite having one to several CCSs that matched the target taxon. For error correction, DADA2 is trained on a pool of sequences and uses sequence abundance to discriminate between sequencing error and true sequence variation. In our case the sequence pools corresponded to each of the 384 samples and were rather small due to the high level of multiplexing (average of 229 CCSs per sample/pool). In addition, among the CCSs available for each pool, in particular for the samples for which DADA2 did not recover the target taxon, the target CCSs were in low abundance within a large pool of lichen-associated fungal sequences or contaminant sequences. Because DADA2 error correction is based on sequence abundance, sequence variants are only inferred for high-abundance sequences. It is therefore not fully applicable to the metabarcoding of lichen herbarium specimens, or at least not when sequences of associated fungi are abundant. In this case, denoising methods that are not based on sequence abundance may perform better.

In terms of sequencing efficiency, with an average sequencing success rate of 86.5%, the new M13 amplicon sequencing protocol from PacBio is comparable to the protocol used in Gueidan et al. (2019). More recently, a UMI-based protocol for both Nanopore and PacBio long-read sequencing, which further decreases chimera and error rates, has been developed (Karst et al. 2021) and would be worth testing as well in

the future. In terms of time, both plate DNA extractions and batch sequence editing allowed us to decrease the time necessary to obtain the final ITS barcode sequences. The method also allows eliminating the time-consuming cloning step for samples with co-amplified fungal products. In terms of cost, the present method allowed us to decrease the cost per sample from AU$37 (Gueidan et al. 2019) to AU$27 (this study). The cost reduction is less than what was anticipated in Gueidan et al. (2019), mostly due to the transition from the RSII sequencing platform to the Sequel sequencing platform, which is more expensive. It is also due to the initial cost of a plate of M13 PacBio universal barcoded primers, which was more expensive than the Barcoded Universal F/R Primers Plate-96 from PacBio used in Gueidan et al. (2019). The M13 PacBio universal barcoded primer plate, however, includes larger volumes and can be used for more reactions. Similarly, although more expensive, the Sequel platform generates a much larger data output than the RSII platform (5–11 Gb versus 500 Mb, respectively). Sequel II, the next generation of SMRT sequencing platform, generates an even larger output (about 80 Gb), while the RSII platform is currently being discontinued in most sequencing services. These changes to the SMRT sequencing platforms and protocols imply that, at present, for lichen specimen metabarcoding, only pooled samples of large numbers of specimens (>500) or larger number of markers will make this technology cost-efficient.

Conclusion

With an average sequencing success of 86.5%, this long-read amplicon sequencing method is confirmed as a potential alternative to Sanger sequencing for the generation of full-length and high-quality DNA barcodes from mixed DNA samples extracted from lichen specimens. It performed particularly well for crustose saxicolous (93.3% success) and squamulose terricolous (96.5% success) taxa. In terms of cost (AU$27/sample), although still more expensive than Sanger sequencing, it allows recovering high-quality sequences even when other lichen-associated fungi amplify as well, eliminating the need for using gel separation or cloning. At high multiplexing level (more than 500 samples/run), this high-throughput method is therefore an attractive option for the generation of DNA barcodes from large number of herbarium specimens.

Acknowledgements

The authors would like to thank Judith Curnow (CANB, Canberra) for her help with specimen databasing and curation, and the BM and NSW herbaria as well as the following collectors and herbaria for providing specimens used in this study: A. Aptroot (ABL), M. Bertrand, J.A. Elix (CANB), G. Kantvilas (HO), M. Mallen-Cooper (UNSW), P.M. McCarthy, B. McCune, R.W. Purdie (CANB), C. Roux (MARSSJ) and D. Stone. They would also like to acknowledge the contribution of Cameron

Jack (ANU, Canberra) and Nunzio Knerr (CANB, Canberra) to various bioinformatic scripts. They also thank Tonia Russell (Ramaciotti, UNSW, Sydney) and Cheryl Heiner (PacBio, USA) for their advice with sample preparation and sequencing.

References

Abarenkov K, Nilsson RH, Larsson KH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E, Pennanen T, Sen R, Taylor AFS, Tedersoo L, Ursing BM, Vrålstad T, Liimatainen K, Peintner U, Kõljalg U (2010) The UNITE database for molecular identification of fungi - recent updates and future perspectives. New Phytologist 186: 281–285. https://doi.org/10.1111/j.1469-8137.2009.03160.x

Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R (2017) Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems 2: e00191-16. https://doi.org/10.1128/mSystems.00191-16

Bendiksby M, Mazzoni S, Jørgensen MH, Halvorsen R, Holien H (2014) Combining genetic analyses of archived specimens with distribution modelling to explain the anomalous distribution of the rare lichen *Staurolemma omphalarioides*: long-distance dispersal or vicariance? Journal of Biogeography 41: 2020–2031. https://doi.org/10.1111/jbi.12347

Blaalid R, Carlsen T, Kumar S, Halvorsen R, Ugland KI, Fontana G, Kauserud H (2012) Changes in the root-associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing. Molecular Ecology 21: 1897–1908. https://doi.org/10.1111/j.1365-294X.2011.05214.x

Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kauserud H (2013) ITS1 versus ITS2 as DNA metabarcodes for fungi. Molecular Ecology Resources 13: 218–224. https://doi.org/10.1111/1755-0998.12065

Bowman BN, Marks P, Hepler NL, Eng K, Harting J, Shiina T, Suzuki S, Ranade S (2014) Long amplicon analysis: highly accurate, full-length, phased, allele-resolved gene sequences from multiplexed SMRT sequencing data. Pacific Biosciences (poster https://www.pacb.com/proceedings/long-amplicon-analysis-highly-accurate-full-length-phased-allele-resolved-gene-sequences-from-multiplexed-smrt-sequencing-data).

Buée M, Reich M, Murat C, Nilsson RH, Uroz S, Martin F (2009) 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. New Phytologist 184: 449–456. https://doi.org/10.1111/j.1469-8137.2009.03003.x

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: high-resolution sample inference from Illumina amplicon data. Nature Methods 13: 581–583. https://doi.org/10.1038/nmeth.3869

Castaño C, Berlin A, Brandström Durling M, Ihrmark K, Lindahl BD, Stenlid J, Clemmensen KE, Olson Å (2020) Optimized metabarcoding with Pacific Biosciences enables semi-quantitative analysis of fungal communities. New Phytologist 228: 1149–1158. https://doi.org/10.1111/nph.16731

Chen Y, Frazzitta AE, Litvintseva AP, Fang C, Mitchell TG, Springer DJ, Ding Y, Yuan G, Perfect JR (2015) Next generation multilocus sequence typing (NGMLST) and the analytical software program MLSTEZ enable efficient, cost-effective, high-throughput, multilocus sequence typing. Fungal Genetics and Biology 75: 64–71. https://doi.org/10.1016/j.fgb.2015.01.005

Cline LC, Zak DR (2015) Initial colonization, community assembly and ecosystem function: fungal colonist traits and litter biochemistry mediate decay rate. Molecular Ecology 24: 5045–5058. https://doi.org/10.1111/mec.13361

Crous PW, Giraldo A, Hawksworth DL, Robert V, Kirk PM, Guarro J, Robbertse B, Schoch CL, Damm U, Trakunyingcharoen T, Groenwald JZ (2014) The Genera of Fungi: fixing the application of type species of generic names. IMA Fungus 5: 141–160. https://doi.org/10.5598/imafungus.2014.05.01.14

Del Prado R, Divakar PK, Crespo A (2011) Using genetic distances in addition to ITS molecular phylogeny to identify potential species in the *Parmotrema reticulatum* complex: a case study. The Lichenologist 43: 569–583. https://doi.org/10.1017/S0024282911000582

Divakar PK, Blanco O, Hawskworth DL, Crespo A (2005) Molecular phylogenetic studies on the *Parmotrema reticulatum* (syn. *Rimelia reticulata*) complex, including the confirmation of *P. pseudoreticulatum* as a distinct species. The Lichenologist 37: 55–65. https://doi.org/10.1017/S0024282904014586

Edgar RC (2016) UNOISE2: Improved error-correction for Illumina 16S and ITS amplicon reads. http://dx.doi.org/10.1101/081257

Gardes M, Bruns TD (1993) ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. Molecular Ecology 2: 113–118. https://doi.org/10.1111/j.1365-294X.1993.tb00005.x

Geml J, Gravendeel B, van der Gaag KJ, Neilen M, Lammers Y, Raes N, Semenova TA, de Knijff P, Noordeloos ME (2014) The Contribution of DNA metabarcoding to fungal conservation: diversity assessment, habitat partitioning and mapping red-listed fungi in protected coastal *Salix repens* communities in the Netherlands. PLoS ONE 9: e99852. https://doi.org/10.1371/journal.pone.0099852

Gueidan C, Aptroot A, Cáceres MES (2016) Molecular phylogeny of the tropical family Pyrenulaceae: contribution from dried specimens and FTA card samples. Mycological Progress 15: 1–21. https://doi.org/10.1007/s11557-015-1154-8

Gueidan C, Elix JA, McCarthy PM, Roux C, Mallen-Cooper M, Kantvilas G (2019) PacBio amplicon sequencing for metabarcoding of mixed DNA samples from lichen herbarium specimens. MycoKeys 53: 73–91. https://doi.org/10.3897/mycokeys.53.34761

Heeger F, Bourne EC, Baschien C, Yurkov A, Bunk B, Spröer C, Overmann J, Mazzoni CJ, Monaghan MT (2018) Long-read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments. Molecular Ecology Resources 18: 1500–1514. https://doi.org/10.1111/1755-0998.12937

Hodkinson BP, Lendemer JC (2013) Next-generation sequencing reveals sterile crustose lichen phylogeny. Mycosphere 4: 1028–1039. https://doi.org/10.5943/mycosphere/4/6/1

Hu Y, Green G, Milgate A, Stone E, Rathjen J, Schwessinger B (2019) Pathogen detection and microbiome analysis of infected wheat using a portable DNA sequencer. Phytobiomes Journal 3: 92–101. https://doi.org/10.1094/PBIOMES-01-19-0004-R

Irinyi L, Serena C, Garcia-Hermoso D, Arabatzis M, Desnos-Ollivier M, Vu D, Cardinali G, Arthur I, Normand A-C, Giraldo A, da Cunha KC, Sandoval-Denis M, Hendrickx M, Nishikaku AS, de Azevedo Melo AS, Merseguel KB, Khan A, Parente-Rocha JA, Sampaio P, da Silva Briones MR, e Ferreira RC, Muniz M de M, Castanon-Olivares LR, Estrada-Barcenas D, Cassagne C, Mary C, Duan SY, Kong F, Sun AY, Zeng X, Zhao Z, Gantois N, Botterel F, Robbertse B, Schoch CL, Gams W, Ellis D, Halliday C, Chen S, Sorrell TC, Piarroux R, Colombo AL, Pais C, de Hoog S, Zancopé-Oliveira RM, Taylor ML, Toriello C, de Almeida Soares CM, Delhaes L, Stubbe D, Dromer F, Ranque S, Guarro J, Cano-Lira JF, Robert V, Velegraki A, Meyer W (2015) International Society of Human and Animal Mycology (ISHAM)-ITS reference DNA barcoding database - the quality controlled standard tool for routine identification of human and animal pathogenic fungi. Medical Mycology 53: 313–337. https://doi.org/10.1093/mmy/myv008

James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE, Hosaka K, Sung GH, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC, Wang Z, Wilson AW, Schüssler A, Longcore JE, O'Donnell K, Mozley-Standridge S, Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies DR, Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D, Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkmann-Kohlmeyer B, Spotts RA, Serdani M, Crous PW, Hughes KW, Matsuura K, Langer E, Langer G, Untereiner WA, Lücking R, Büdel B, Geiser DM, Aptroot A, Diederich P, Schmitt I, Schultz M, Yahr R, Hibbett DS, Lutzoni F, McLaughlin DJ, Spatafora JW, Vilgalys R (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature 443: 818–22. https://doi.org/10.1038/nature05110

James TY, Marino JA, Perfecto I, Vandermeer J (2016) Identification of putative coffee rust mycoparasites via single-molecule DNA sequencing of infected pustules. Applied Environmental Microbiology 82: 631–639. https://doi.org/10.1128/AEM.02639-15

Karst SM, Ziels RM, Kirkegaard RH, Sørensen EA, McDonald D, Zhu Q, Knight R, Albertsen M (2021) High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. Nature Methods 18: 165–169. https://doi.org/10.1038/s41592-020-01041-y

Kelly LJ, Hollingsworth PM, Coppins BJ, Ellis CJ, Harrold P, Tosh J, Yahr R (2011) DNA barcoding of lichenized fungi demonstrates high identification success in a floristic context. New Phytologist 191: 288–300. https://doi.org/10.1111/j.1469-8137.2011.03677.x

Kistenich S, Halvorsen R, Schrøder-Nielsen A, Thorbek L, Timdal E, Bendiksby M (2019) DNA sequencing historical lichen specimens. Frontiers in Ecology and Evolution 7: e5. https://doi.org/10.3389/fevo.2019.00005

Kõjalg U, Larsson K-H, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E, Pennanen T, Sen R, Taylor AFS, Tedersoo L, Vrålstad T, Ursing BM (2005) UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytologist 166: 1063–1068. https://doi.org/10.1111/j.1469-8137.2005.01376.x

Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AF, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, Douglas B, Drenkhan T, Eberhardt U, Dueñas

M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lücking R, Martín MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Põldmaa K, Saag L, Saar I, Schüßler A, Scott JA, Senés C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M, Larsson KH (2013) Towards a unified paradigm for sequence-based identification of fungi. Molecular Ecology 22: 5271–5277. https://doi.org/10.1111/mec.12481

Larrea A, Yuan G, Wagner K, Butt B, Norvell B, Aro L, Gu J, Lleras R, Ranade S, Korlach J, Krueger B, Heiner C (2018) A simple segue from Sanger to high-throughput SMRT Sequencing with a M13 barcoding system. Pacific Biosciences (poster https://www.pacb.com/wp-content/uploads/Aro-ASHG-2018-Targeted-Sequencing-M13-10-6-2018-FINAL.pdf).

Lohtander K, Myllys L, Sundin R, Källersjö M, Tehler A (1998) The species pair concept in the lichen *Dendrographa leucophaea* (Arthoniales): analyses based on ITS sequences. Bryologist 101: 404–411. https://doi.org/10.2307/3244179

Loit K, Adamson K, Bahram M, Puusepp R, Anslan S, Kiiker R, Drenkhan R, Tedersoo L. (2019) Relative performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) third-generation sequencing instruments in identification of agricultural and forest fungal pathogens. Applied Environmental Microbiology 85: e01368–19. https://doi.org/10.1128/AEM.01368-19

Lücking R, Lawrey JD, Gillevet PM, Sikaroodi M, Dal-Forno M, Berger SA (2014) Multiple ITS haplotypes in the genome of the lichenized basidiomycete *Cora inversa* (Hygrophoraceae): Fact or artifact? Journal of Molecular Evolution 78: 148–162. https://doi.org/10.1007/s00239-013-9603-y

Lumini E, Orgiazzi A, Borriello R, Bonfante P, Bianciotto V (2010) Disclosing arbuscular mycorrhizal fungal biodiversity in soil through a land-use gradient using a pyrosequencing approach. Environmental Microbiology 12: 2165–2179. https://doi.org/10.1111/j.1462-2920.2009.02099.x

Mark K, Cornejo C, Keller C, Flück D, Scheidegger C (2016) Barcoding lichen-forming fungi using 454 pyrosequencing is challenged by artifactual and biological sequence variation. Genome 59: 685–704. https://doi.org/10.1139/gen-2015-0189

Mello A, Napoli C, Morin CME, Marceddu G, Bonfante P (2011) ITS-1 versus ITS-2 pyrosequencing: a comparison of fungal populations in truffle grounds. Mycologia 103: 1184–1193. https://doi.org/10.3852/11-027

Myllys L, Lohtander K, Källersjö M, Tehler A (1999) Sequence insertions and ITS data provide congruent information on *Roccella canariensis* and *R. tuberculata* (Arthoniales, Euascomycetes) phylogeny. Molecular Phylogenetics and Evolution 12: 295–309. https://doi.org/10.1006/mpev.1999.0620

Nilsson RH, Veldre V, Hartmann M, Unterseher M, Amend A, Bergsten J, Kristiansson E, Ryberg M, Jumpponen A, Abarenkov K (2010) An open-source package for automated extraction of ITS1 and ITS2 from fungal ITS sequences for use in high-throughput community assays and molecular ecology. Fungal Ecology 3: 284–287. https://doi.org/10.1016/j.funeco.2010.05.002

Nilsson RH, Taylor AFS, Adams RI, Baschien C, Bengtsson-Palme J, Cangren P, Coleine C, Daniel H-M, Glassman SI, Hirooka Y, Irinyi L, Iršėnaitė R, Martin-Sanchez PM, Meyer

W, Oh S-Y, Sampaio JP, Seifert KA, Sklenař F, Stubbe D, Suh S-O, Summerbell R, Svantesson S, Unterseher M, Visagie CM, Weiss M, Woudenberg JHC, Wurzbacher C, Van den Wyngaert S, Yilmaz N, Yurkov A, Kõljalg U, Abarenkov K (2018) Taxonomic annotation of public fungal ITS sequences from the built environment – a report from an April 10–11, 2017 workshop (Aberdeen, UK). MycoKeys 28: 65–82. https://doi.org/10.3897/mycokeys.28.20887

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD (2015) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research 44: 733–745. https://doi.org/10.1093/nar/gkv1189

Orock EA, Leavitt SD, Fonge BA, St Clair LL, Lumbsch HT (2012) DNA-based identification of lichen-forming fungi: can publicly available sequence databases aid in lichen diversity inventories of Mount Cameroon (West Africa)? The Lichenologist 44: 833–839. https://doi.org/10.1017/S0024282912000424

Redchenko O, Vondrák J, Košnar J (2012) The oldest sequenced fungal herbarium sample. The Lichenologist 44: 715–718. https://doi.org/10.1017/S002428291200031X

Schlaeppi K, Bender SF, Mascher F, Russo G, Patrignani A, Camenzind T, Hempel S, Rillig MC, van der Heijden MGA (2016) High-resolution community profiling of arbuscular mycorrhizal fungi. New Phytologist 212: 780–791. https://doi.org/10.1111/nph.14070

Schoch CL, Seifert KA, Huhndorf C, Robert V, Spouge JL, Levesque CA, Chen W, and Fungal Barcoding Consortium (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proceedings of the National Academy of Sciences USA 109: 6241–6246. https://doi.org/10.1073/pnas.1117018109

Schoch CL, Robbertse B, Robert V, Vu D, Cardinali G, Irinyi L, Meyer W, Nilsson RH, Hughes K, Miller AN, Kirk PM, Abarenkov K, Aime MC, Ariyawansa HA, Bidartondo M, Boekhout T, Buyck B, Cai Q, Chen J, Crespo A, Crous PW, Damm U, De Beer ZW, Dentinger BTM, Divakar PK, Dueñas M, Feau N, Fliegerova K, García MA, Ge Z-W, Griffith GW, Groenewald JZ, Groenewald M, Grube M, Gryzenhout M, Gueidan C, Guo L, Hambleton S, Hamelin R, Hansen K, Hofstetter V, Hong S-B, Houbraken J, Hyde KD, Inderbitzin P, Johnston PR, Karunarathna SC, Kõljalg U, Kovács GM, Kraichak E, Krizsan K, Kurtzman CP, Larsson K-H, Leavitt S, Letcher PM, Liimatainen K, Liu J-K, Lodge DJ, Luangsa-Ard JJ, Lumbsch HT, Maharachchikumbura SSN, Manamgoda D, Martín MP, Minnis AM, Moncalvo J-M, Mulè G, Nakasone KK, Niskanen T, Olariaga I, Papp T, Petkovits T, Pino-Bodas R, Powell MJ, Raja HA, Redecker D, Sarmiento-Ramirez JM, Seifert KA, Shrestha B, Stenroos S, Stielow B, Suh S-O, Tanaka K, Tedersoo L, Telleria MT, Udayanga D, Untereiner WA, Uribeondo JD, Subbarao KV, Vágvölgyi C, Visagie C, Voigt K, Walker DM, Weir BS, Weiß M, Wijayawardene NN, Wingfield MJ, Xu JP, Yang ZL,

Zhang N, Zhuang W-Y, Federhen S (2014) Finding needles in haystacks: linking scientific names, reference specimens and molecular data for Fungi. Database 2014: 1–21. https://doi.org/10.1093/database/bau061

Sohrabi M, Myllys L, Stenroos S (2010) Successful DNA sequencing of a 75 year-old herbarium specimen of *Aspicilia aschabadensis* (J. Steiner) Mereschk. Lichenologist 42: 626–628. https://doi.org/10.1017/S0024282910000344

Staiger B, Kalb K, Grube M (2006) Phylogeny and phenotypic variation in the lichen family Graphidaceae (Ostropomycetidae, Ascomycota). Mycological Research 110: 765–772. https://doi.org/10.1017/S0024282910000344

Tedersoo L, Anslan S (2019) Towards PacBio-based pan-eukaryote metabarcoding using full-length ITS sequences. Environmental Microbiology Reports 11: 659–668. https://doi.org/10.1111/1758-2229.12776

Tzovaras BG, Segers FHID, Bicker A, Dal Grande F, Otte J, Anvar SY, Hankeln T, Schmitt I, Ebersberger I (2020) What is in *Umbilicaria pustulata*? A metagenomic approach to reconstruct the holo-genome of a lichen. Genome Biology and Evolution 12: 309–324. https://doi.org/10.1093/gbe/evaa049

Walder F, Schlaeppi K, Wittwer R, Held AY, Vogelgsang S, van der Heijden MGA (2017) Community profiling of *Fusarium* in combination with other plant-associated fungi in different crop species using SMRT sequencing. Frontiers in Plant Science 8: e2019. https://doi.org/10.3389/fpls.2017.02019

Weerakoon G, Aptroot A, Lumbsch HT, Wolseley PA, Wijeyaratne SC, Gueidan C (2012) New molecular data on Pyrenulaceae from Sri Lanka reveal two well supported groups within this family. The Lichenologist 44: 639–647. https://doi.org/10.1017/S0024282912000333

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (Eds) PCR Protocols, a Guide to Methods and Applications. Academic Press, San Diego, 315–322. https://doi.org/10.1016/B978-0-12-372180-8.50042-1

Yahr R, Schoch C, Dentinger BTM (2016) Scaling up discovery of hidden diversity in fungi: impacts of barcoding approaches. Philosophical Transactions of the Royal Society B 371: e20150336. https://doi.org/10.1098/rstb.2015.0336

Supplementary material I

Table S1. List of specimens used for this study, including their voucher information, plate location, indexing, amplicon concentration and sequencing results, both as an output from SMRT tools (CCSs) and as an output from DADA2 (sequence variants). Table S2. List of the 64 barcode sequences used to index the samples. Used barcode pairs are listed in Table S1

Authors: Cécile Gueidan, Lan Li

Data type: Taxon sampling

Explanation note: List of specimens used for this study, including their voucher information, plate location, indexing, amplicon concentration and sequencing results, both as an output from SMRT tools (CCSs) and as an output from DADA2 (sequence variants). A summary of the blast results for the sequence variants is also listed for each sample. In the "recovered target" column, samples for which the target sequence was recovered from the CCS file but not the sequence variant file are indicated by a star.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mycokeys.86.77431.suppl1